

Title: Leveraging non-lattice fragments for QA purposes

Presenter: Licong Cui^a PhD, Olivier Bodenreider^b MD, GQ Zhang^a PhD
a: Case Western Reserve University, USA
b: The National Library of Medicine, USA

Audience

SNOMED CT developers and users who are interested in continued quality improvement and assurance work.

Objectives

1. To learn about lattice vs. non-lattice fragments in the hierarchical structure of SNOMED CT;
2. To learn how non-lattice fragments often reveal missing or misaligned hierarchical relations;
3. To assess the potential of non-lattice fragments for quality assurance in SNOMED CT.

Abstract

In the SNOMED CT inferred hierarchy, “Irritable bowel syndrome variant of childhood” (235838003) and “Irritable bowel syndrome with diarrhea” (197125005) have both “Irritable bowel syndrome” (10743008) and “Disorder of colon” (128524007) as shared parents, and “Irritable bowel syndrome” is not classified as a “Disorder of colon,” as it should (missing hierarchical relation). This configuration of hierarchical relations in which two concepts have multiple shared parents is characteristic of “non-lattice” fragments. If “Irritable bowel syndrome” was a child of “Disorder of colon,” “Irritable bowel syndrome variant of childhood” and “Irritable bowel syndrome with diarrhea” would only have “Irritable bowel syndrome” as a single shared parent, and these two concepts would form a lattice fragment. We show that identifying and analyzing non-lattice fragments in the SNOMED CT hierarchy is a simple and efficient quality assurance technique.

The rationale for analyzing non-lattice fragments is established in [1]. A computationally scalable pipeline using Hadoop for extracting all non-lattice fragments in SNOMED CT is introduced in [2]. (In addition to the simple example above, non-lattice fragments may involve shared ancestors beyond direct parents, which makes their identification a non-trivial, computationally intensive task.)

This presentation focuses on analyzing non-lattice fragments for missing or misaligned hierarchical relations. We identified a total of 160,322 and 165,694 non-lattice fragments in the 2014/09 and 2015/03 versions of SNOMED CT (US edition), respectively. We evaluated 50 smallest (4-node) non-lattice fragments in both versions (one fragment in the 2014/09 version has been corrected in the 2015/03 version). The example above is one such fragment. We provide analysis and clustering of error-types on the 49 4-node fragments to gain insights into a powerful approach for quality assurance (QA) in SNOMED CT, complementary to description logics-based QA.

References

1. Zhang, GQ. and O. Bodenreider, Large-scale, exhaustive lattice-based structural auditing of SNOMED CT. AMIA Annu Symp Proc, 2010: p. 922-926.
2. Zhang GQ, Zhu W, Sun M, Tao S, Bodenreider O, Cui L. MaPLE: A MapReduce Pipeline for Lattice-based Evaluation and Its Application to SNOMED CT. IEEE International Conference on Big Data, 2014;754-759.